# VA Summer Epidemiology Session
# Developing Scientific Research Proposals (Grant Writing)

## **Session 7** (Readings)

### Methods: Analysis and Power

**A. Guidelines**
**B. Case Control Study**
**C. Cohort Study**
**D. Intervention Trial**
**E. Ancillary Study to Intervention Trial**
**F. Power**
**F.1. Overview**
**F.2. By Study Type**

---

**A. Guidelines**

The Analysis section describes exactly how you will test the null hypotheses given in your Specific Aims. The Power section describes relationships among four parameters: (1) effect size; (2) sample size; (3) probability of a statistically significant test when $H_0$ should be rejected (1-β or power); and (4) probability of a statistically significant test when $H_0$ should not be rejected (Type I or α error). Because type I error is usually set to a constant (most often 0.05), results of power analyses are usually expressed as a function of any two of the three remaining parameters. At study section, every epidemiologist and biostatistician will go directly to your Analysis and Power sections. Your analyses must be clear, the statistical test must be appropriate, and the power analyses must show that your design will likely lead to a meaningful result.

Here are some guidelines for the analysis section.

- Link each analysis directly and explicitly to a specific aims.
- Specify each independent and dependent variable and covariate.
- Specify the statistical model, procedure or test
- Describe how each variable is coded for analysis (continuous, transformed, categorized, etc.)
- Justify each decision
- Specify precisely which parameter(s) is (are) used to test the specific aim

Examine the Analysis sections below.

---

## B. For a Case Control Study

### Analyses of Specific Aims

Primary Aim: Association of Dietary Fat Intake with Prostate Cancer. As an example of our analytic approach, we present the analyses for the first part of the primary specific aim: the investigation of the association of total dietary fat intake with prostate cancer.

Choice of a Statistical Model. We will analyze the association between dietary fat intake and biopsy proven presence or absence of prostate cancer using the logistic regression model:

$$ln\left(\frac{pr(d)}{1-pr(d)}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots.$$

The logistic model allows estimation of the effect of the factor of interest ($x_1$), in this case fat intake, on probability of disease (pr(d)), while controlling the effects of other factors ($x_2, x_3 \ldots$). We plan to control for age, race, family history, and other confounding factors identified during analysis. Parameters in the model will be estimated using maximum likelihood techniques.[102]

Although this is a cohort study, survival analysis is inappropriate. Our endpoints are not time-to-event; rather they are simple presence or absence of prostate cancer at seven years. We assume that any persons diagnosed with prostate cancer before seven years would also be positive for prostate cancer at seven years.

Expression of Fat Intake in the Model and Adjustment for Energy. There are several approaches to expressing fat intake and adjusting for energy intake in epidemiologic models.[90] Adjustment of fat intake for total energy intake is important, especially when the dietary assessment tool is a food frequency questionnaire, because some people tend to overestimate and others underestimate their total energy intake. In addition, there are individual differences in total energy intake due to differences in body size, physical activity or metabolic factors, and it may be desirable to adjust for these factors. However, because fat intake and energy intake are highly correlated (r = .88)[90], statistical control of fat intake for total energy is complex. There is considerable interest and research on the best ways to model dietary intake data in epidemiologic analyses, and we expect considerable progress before the data from this study will be analyzed. We present here the two approaches that now appear best to us.

First, we will express fat intake as total energy from fat and we will adjust for total energy from other macronutrients (carbohydrate, protein and alcohol). Second, we will express fat intake as percent of energy from fat:

$$\text{\% energy from fat} = \frac{\text{fat intake (g) x 9 kcals / g}}{\text{total energy intake (kcal)}}.$$

Percent energy from fat is an attractive way of expressing fat intake both from a biologic perspective (it expresses the nutrient density of food intake) and from a public health approach (most public health recommendations concerning fat intake are given in terms of percent energy from fat). This expression for fat intake has a statistical advantage as well, because it is only weakly correlated with energy intake (r = .05).[90] We will also include energy intake as an adjustment factor in this model, which will adjust for any remaining confounding.

We will model the association between fat intake and prostate cancer by categorizing the fat measure into quartiles (using indicator variables in the model), and we will calculate the odds ratios of prostate cancer and their 95% confidence intervals for each quartile of fat intake. This

model is commonly used in epidemiology, because it allows interpretation of the shape and magnitude of the association of the factor of interest with disease risk. We will also conduct a test for trend by representing percent energy from fat as a single variable coded 1 for quartile 1, coded 2 for quartile 2, etc., in the logistic model. This yields an overall estimate of the trend odds ratio for a one quartile increment in fat intake and yields a single significance test for the association of fat intake with prostate cancer. This is the logistic analog to the Mantel-Haenszel test for trend.

For most men, we will use the average of values for the one- and four-year FFQs to improve the accuracy of the measurement of fat intake.[85] For men who only complete one, only one will be used. For men who have a biopsy (positive or negative) before the second FFQ, only the first will be used. Because averaging two measures reduces the random component of measurement error[85], inclusion of some subjects with only one measure should not introduce any systematic difference in measurement error between cases and controls. We will examine results using only the first FFQ, to assure us that there is no bias.

The other variables to be analyzed as part of the primary specific aim are saturated fat intake and servings of high fat meats and servings of high fat dairy foods. The relation of saturated fat to prostate cancer will be analyzed as described for total fat intake. The other variables will modeled similarly, by categorizing servings per day into quartiles and adjusting for total energy.

Secondary Aims A and B: Associations of Fiber and Micronutrients with Prostate Cancer. The variables to be analyzed as part of the secondary aim are dietary fiber, β-carotene, vitamin A, servings of fruits and servings of vegetables. We will also analyze intake of other micronutrients (Table 3). In these analyses, we will categorize fiber, micronutrients and servings of food groups into quartiles, and we will adjust for total energy. Adjustment of total energy is methodologically more straightforward in these models, because correlations between these dietary measures and total energy are modest (e.g., the correlation between fiber and total energy is approximately 0.35). However, the best approach to analyses of micronutrients that includes intake from nutritional supplements is unclear. Variability in micronutrient intake due to use of supplements often overwhelms any variability in micronutrient intake from foods, and micronutrients from multivitamin supplements are so strongly correlated that it may not be possible to separate effects of specific micronutrients. We will therefore also analyze micronutrients: (1) from foods alone; (2) from supplements alone; and (3) from foods and supplements together. We will also examine the effects of using multivitamins, coded as a bivariate indicator variable, both as an independent variable and as a covariate.

Secondary Aim D: Modification of the Effect of Finasteride by Dietary Fat. The final analyses for this study will be conducted to answer whether the effect of finasteride is modified by dietary fat intake. First, we will compute the odds ratios and 95% confidence intervals for the effect of finasteride on prostate cancer separately for the four groups of men categorized by quartiles of percent energy from fat. If these appear to differ (e.g., if finasteride is beneficial for men with the highest level of fat intake but not for those in the lowest level of intake), this is suggestive of modification of the finasteride effect by dietary fat intake.

Second, we will model the interaction between finasteride and dietary fat. To examine this interaction, we will model the relationship between dietary fat, treatment and prostate cancer in a logistic model using the entire study cohort. The model will include the main effect of treatment (finasteride vs. control) and the main effects of dietary fat and an interaction between the two, as well as covariates (age, race, energy intake). If consistent with the data, dietary fat will be modeled as a single trend variable (described under Primary Aim above). This simplifies the interaction, so that a single term would indicate whether the slope of any dietary fat effect on prostate cancer is different for the treatment vs. the control group. We will use the interaction term coefficient (β) and its standard error to determine the magnitude and significance of the modification of the finasteride effect by dietary fat.

## C. For a Cohort Study

### DATA ANALYSIS AND STUDY POWER

**Outcomes.** The primary outcome is total cancer incidence. Although both in situ and invasive cancers will be ascertained, the primary analysis will be limited to invasive cancer.

**Definition of exposure variables.** The four primary exposure variables are average intake per day of supplemental vitamin C, vitamin E, calcium, and multivitamins over the 10 year reference period ending at baseline (Aim 2). Cumulative intake of supplemental vitamin E, C, and calcium will be computed by summing intake from individual supplements of that nutrient (based on years of use, days per week used in those years, and reported dose per day) and intake of that nutrient from multivitamins in the 10 year period (based on years of use of each type of multivitamin, days used per week in those years, and actual formulation of the named brand or if not available, the most common formulation of that type of multivitamin—see Section D.3). This yields the 10 year cumulative dose, which is converted to average dose per day over the entire 10 year period for ease of interpretation.

For multivitamin pills, cumulative dose will simply be expressed as years of daily use (years x days per week used in those years/7). The primary analysis will be restricted to multivitamin pills with minerals. This is the most common formulation, and interpretation of the results will be straightforward if only one type is included. We will, however, also use other categories of multivitamins, e.g., any type of multivitamin pill with retinol.

These supplement variables will be divided into four categories: no use over the 10 year period and tertiles of use among users. The 33$^{rd}$ and 67$^{th}$ percentile of average daily dose of supplements over the 10 year period among users from out pilot study is given in Table 5; these are estimates of the cutpoints for the categories of users. Each of the cutpoints for the highest third, except calcium use among men, is significantly above what could be achieved by daily use of a multivitamin pill. As noted in Section B Limitations of Past Studies, collinearity of nutrients from supplements induced by taking multivitamin supplements has been a problem in past studies. However, we will have sufficient numbers of subjects taking individual supplements of the micronutrients of interest (see Table 5), so that only users of individual supplements would fall into the highest third of intake. We prefer the categorization of variables because it provides risk estimates for each of several levels of use, rather than attempting to model the dose-response relationship with one parameter. However, other parameterizations of these variables will also be considered.

In addition to these primary exposure variables, we will also consider separately the average dose per day in the years the supplement was taken and the duration of use, to understand how these two components (duration and intensity) of cumulative dose affect the outcomes.

Our secondary exposures (Aim 3) are intakes of total (diet plus supplementary) vitamin C, E, and calcium. These will be estimated by summing average daily intake of these nutrients from the FFQ and from the supplement questionnaire, based on supplements currently taken at baseline. Attention will be paid to the forms of the nutrient in supplements and in food. In

particular, vitamin A in multivitamin pills now includes both vitamin A acetate and β-carotene, and vitamin E in supplements is dl-α tocopheryl acetate, while food tables reflect the α tocopherol content of food. We will use weights to standardize the potencies in terms of retinol equivalents for vitamin A or α tocopherol equivalents for vitamin E (129) when different forms are added together.

**Confounding and effect modification.** Supplement users may be more health conscious and therefore more likely to practice other disease prevention behaviors, which may reduce their risk of disease (14). Our pilot study and others (13, 15, 130) have shown some differences in smoking, exercise, diet and screening between supplement users and non-users. We will use two methods to control for factors that may distinguish supplement users from non-users. First, dietary intake, intake of other supplements, and the other health behaviors listed in Table 8 will be tested as potential confounders. Of particular importance are the need to test as possible confounders nutrient intake variables, e.g., control for vitamin C intake from diet when considering vitamin C from supplements and control for other supplement use, e.g., use of vitamin E. The second method is to look for a dose-response gradient of supplement use on the study outcome only among users of that supplement (e.g., a trend across the three levels of use of vitamin C). Thus, only users of supplemental vitamin C are compared among themselves, reducing bias caused by comparing non-users to users (confounding factors would of course still be evaluated). Although this approach would have less power, it would add evidence for or against specific effects.

Conversely, there is also concern that certain health conditions may lead people to take supplements and these health conditions would increase the outcome events in supplement users. As noted under Section D.3 above, this concern is greater for mortality than for cancer incidence. To reduce the problem that those with serious medical conditions might have begun using supplements, we will exclude the first two years of deaths from analyses of the relation between supplements and mortality. (Therefore mortality is not an endpoint in this 5-year proposal.) For cancer, we will use two methods to control for this type of confounding. First, we will stratify our analyses by the major risk factors and preclinical conditions that could have prompted individuals to begin supplement use (e.g., family history, benign breast biopsies), to evaluate any differential effect of these conditions on the relation between supplement use and cancer. If effect modification is not observed, we will combine the groups but control for these and other factors. Secondly, we can control for the major reasons for taking supplements (for a current or prior health condition, to prevent future diseases, for more energy, etc.).

In addition, we will stratify our analyses by sex, by smoking (current, former, never and/or pack-years) and by other potential modifiers, and test for effect modification.

**Statistical model.** For a given supplemental intake variable, we will create indicator variables $x_1, x_2, x_3$ for intake in the three tertiles of use of that supplement. Relative risks and their confidence intervals adjusted for covariates for these three tertiles of use will be estimated from Cox proportional hazard models (131):

$$\ln \lambda(t) = \ln \lambda_0(t) + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \text{ other covariates, [1]}$$

where $\lambda_0(t)$ is the cancer incidence rate among non-users. The test for trend across the four levels of supplement intake will be performed by creating the variable y taking values of 0, 1, 2, and 3 for non-users, and first, second, and third tertile of use, respectively, and testing the parameter β in the model:

$$\ln\ \lambda(t) = \ln \lambda_0(t) + \beta y + \text{other covariates}. \quad [2]$$

For analyses involving data from more than one individual in a household, we will assess the level of intra-household correlation in cancer incidence using methods for bivariate survival analysis (132). We anticipate that, after adjustment for known risk factors and behaviors, intrahousehold correlations will be small and have little effect on parameter estimates and confidence intervals.

Other covariates will be classified as confounders and included in the analyses if they modify the point estimate of $\beta$ in [2] by 10% or more, determined by comparing the estimates with and without the potential confounder. Effect modification of $\beta$ by a categorical predictor z will be performed by the likelihood ratio test between a model with y and z (and other covariates) and a model with y, z, and interaction terms between y and z (and other covariates).

---

## D. For an Intervention Trial

### Analysis and Statistical Power

Our principal analyses will be based on treatment group contrasts in: (Specific Aim 1) changes in %G1 and %S phase between baseline and 36 months; and (Specific Aim 2) percent of participants developing islands of normal epithelium within Barrett's columnar epithelium at 36 months. Secondary analyses will address: (Secondary Aim 1) an additional measure of neoplastic progression, aneuploidy and/or accumulation of cells at the G2 cell cycle checkpoint; and (Secondary Aim 2) relationships between the dietary intervention and gastroesophageal reflux, assessed as symptoms and endoscopic evidence of ulcers, erosions and related lesions.

Principal Outcomes

We will use two markers of neoplastic progression, %G1 and %S phase, as the principal outcome measures for Specific Aim 1. We will use the mean of nine biopsies from each assessment point to characterize values for each participant. We will use a linear regression model: %G1 (month 36) = b0 + b1*I + b2*X, where b0 is the intercept, I is the treatment indicator, and X is the vector of covariates (age, body mass index, and time since Barrett's diagnosis). The statistical test that b1 is significantly larger than zero will address Specific Aim 1. The %S endpoint will be analyzed similarly.

We will use a binary response indicator of whether the participant developed islands of normal squamous epithelium at 36 months as the outcome measure for Specific Aim 2. We will use a logistic regression model: logit (p) = b0 + b1*I + b2*X, where p is the probability of a patient developing squamous islands, b0 is the intercept, I is the treatment indicator, and X is the vector

of covariates (age, body mass index, and time since Barrett's diagnosis). The statistical test that b1 is significantly larger than zero will address Specific Aim 2.

---

## E. Ancillary Study to Intervention Trial

From the SELECT Trial, and 2X2 factorial design testing selenium and/or vitamin E for prevention of prostate cancer. This ancillary study is to test effects of the study agents on respiratory function, using forced expiratory volume in the first second (FEV1) as the outcome measure.

**Respiratory Ancillary Study: Analysis plan**

All analyses are based on the difference between treatment groups in the change in $FEV_1$ between baseline and follow-up, a measure often termed the "intervention effect" and defined as follows:

$$((FEV_1)_b - (FEV_1)_f)_I - ((FEV_1)_b - (FEV_1)_f)_P$$

where the subscripts *b* and *f* refer to baseline and follow-up, and the subscripts *I* and *P* refer to intervention and placebo groups. We will follow the analytic strategy used in the SELECT parent study. Thus, our analysis plan is not based on a 2X2 factorial design because experimental evidence suggests an interaction of selenium and vitamin E. For the Primary Specifics Aims, the analysis plan consists of three pre-specified comparisons in the intervention effect at 3-years post-randomization, based on the contrasts of (1) selenium vs. placebo; (2) selenium plus vitamin E vs. placebo and (3) Vitamin E vs. placebo. The statistical test will be based on a multiple regression analysis, in which the dependent variable will be difference in $FEV_1$ between baseline and the 3-year follow-up, and the independent variables will be $FEV_1$ at baseline and an indicator variable for treatment arm. In this model, the regression coefficient for the indicator variable is the intervention effect, and the standard error of the regression coefficient is used to test whether this difference is statistically significant. Using this same model, we can also control for baseline characteristics that may affect $FEV_1$, including age, height, smoking, and dietary and serum antioxidants. Recognizing that each of our three pair-wise comparisons is correlated due to the common placebo reference group, the multiple comparison procedure proposed by Dunnett will be employed (95). With 4 treatment groups (including placebo), and a two-sided alpha=0.05, the standard Z critical value for significance is 2.35. This translates into a two-sided p-value needing to be $\leq 0.018$ in order to be statistically significant, giving us a conservative test.

For the secondary specific aims, the analysis plan consists of two pre-planned comparisons in the intervention effect, based on contrasts of (1) selenium plus vitamin E vs. vitamin E alone; and (2) selenium plus vitamin E vs. selenium alone. As is suggested by Proschan (96), these latter two comparisons will be carried out only if one or more of the three primary comparisons are statistically significant. These latter two comparisons will be made using a two-sided alpha=0.025.

The last secondary aim, aim 2.3, investigates whether the effect of intervention is modified by either smoking status or by level of urinary excretion of $F_2$-isoprostanes. To examine effect modification by cigarette smoking, the analyses will be extended to include an indicator variable for smoking, and an interaction term (indicator variable for smoking x indicator variable for treatment). In this model, the regression coefficient for the interaction term denotes the difference in the intervention effect between smokers and non-smokers, and the standard error of

this regression coefficient is used to test whether this difference is statistically significant. We will also explore other approaches to analyzing effect modification by smoking. We will consider smoking dose, parameterizing this variable either as cigarettes per day or as an ordered categorical variable for 0, 1-9, 10-19, and 20+ cigarettes per day. Further, we will examine effects of lifetime smoking in addition to current smoking using smoking history questionnaire responses to compute cumulative smoking dose. Questionnaire updates on smoking status allow the identification of those who quit smoking over the follow-up period, and this will be considered in the analysis. We will use a similar approach to analyze effect modification by urinary $F_2$-isoprostanes. The simplest analysis will dichotomize urinary $F_2$-isoprostanes, and the regression coefficient for the interaction of dichotomized urinary $F_2$-isoprostanes with treatment is the difference in treatment effect between participants with low and high urinary $F_2$-isoprostanes levels. We will also examine the distribution of the continuous data for urinary $F_2$-isoprostanes, and after transformation for normalization, analyze its effects and interaction with treatment when coded as a continuous variable.

The length of the study is five years, to allow for a three year follow-up on all participants, given that in the first six months the study protocols will be implemented at field sites, enrollment will take place from six months to 1.5 years, and the last person enrolled at 1.5 years will be followed for three years for a final measurement at 4.5 years on the study timeline. Thus, the analysis is planned for the final 6 months of the project, to allow for a full three years of follow-up on all men.

In a subgroup of 750 participants, pulmonary function testing will be done yearly to identify the pattern of response to treatment. The data to identify pattern of response is important because although various scenarios may yield the same approximate intervention effect at year 3, they have very different implications for further benefits of continued monitoring of the study population. For example, if supplementation brings about an initial improvement in $FEV_1$ without any evidence of a decreased slope in $FEV_1$ over time there may be little or no benefit to continued monitoring of participants. On the other hand, if supplementation changes the slope of decline in $FEV_1$, it would be interesting to continue follow-up to directly assess whether the changes in slope are maintained or strengthened with further supplementation. Regardless of the pattern of response, it would be very interesting to add a pulmonary function assessment of the Respiratory Ancillary Study participants at the last SELECT visit: further funds will be sought for this follow-up. We have planned annual $FEV_1$ assessments on a subgroup of 750 men to allow investigation of the pattern of response to supplementation (i.e., an average of about 190 men in each of four groups). We will use regression spline models [(97); p 392], for example a first power or linear spline model, and non-parametric smoothing methods [(97); p425] to describe the pattern of the change in $FEV_1$ across the three years ($FEV_1$ measured at baseline, year 1, year 2 and year 3) in each of the treatment groups. This approach assumes no underlying form, and lets the data determine the shape of the relation. Previous epidemiologic research has demonstrated the usefulness of such methods for estimating the dose-response relation between an exposure and an outcome (98).

Additional analysis plans: Additional study data are available to pursue secondary questions. For example, one purpose of the dietary assessment measures in the respiratory ancillary study is to investigate whether the associations of experimental supplements, vitamin E and selenium, with pulmonary outcomes differs according to pre-randomization intake of supplemental vitamin E or supplemental selenium. These data also will allow investigation of the dietary intake of Vitamin C at randomization in relation to $FEV_1$ change over follow-up. The data on serum concentration of serum vitamin E and serum selenium will allow investigation of whether the effects of study supplements on pulmonary function are conditional upon the starting levels of these antioxidants. The results of these analyses can be used to better understand the study's primary outcomes, and will be important in formulating public health recommendations for respiratory disease prevention.

## F. Power Section

### F.1. Overview of the Basics

The power section will give results of analyses that answer one of three questions.

1. The sample size needed to obtain a statistically significant results with a specified probability (power), type 1 error, and effect size.
2. The probability that your analyses will yield statistically significant results (power) given a specified effect size, type 1 error, and sample size
3. The minimum effect size that can be detected as statistically significant with a specified probability (power), type 1 error, and sample size

It is critical to understand that "effect size" is the *true*, underlying effect size, which is the point estimate around which your *observed* effect size will vary. One common misperception is that the effect size is the smallest effect size that would be declared statistically significant, given the measure's variance and sample size. The problem with this reasoning is that this calculation does not incorporate the likelihood that, by chance alone, you can observe effect sizes considerable smaller the true effect size. Power is thus the probability that you will have a statistically significant test statistic for your observed effect, given the variance around the true, underlying effect size and the specified sample size.

Number 1 above is the most common form for power analyses, to solve for a required sample size given a specified effect size and probability of calling the effect statistically significant. For cohort or case-control studies, one usually selects 80% power at a 5% alpha error to determine the required number of participants. For clinical trials, one usually specifies 90% power and 5% alpha error to detect the specified effect size. This is because clinical trials are very expensive, and the cost of additional participants is considered offset by the possibility that a true effect may be missed by chance alone.

Number 2 above is used when there are a fixed number of available observations, for example when you are analyzing an existing data set, and when there is some well-justified minimum level of effect that has biological or public health meaning. For example, suppose you had data from a case control study with 200 cases and 400 controls, and you were interested in analyzing stored bloods to determine if a genetic polymorphism was associated with disease risk, and that the minimum level of increased risk that is meaningful is a relative risk of 1.5. Your power calculation would state how much power you would have to detect a relative risk of this magnitude (e.g., 84%). This is also the approach to use when giving power for secondary aims, when the primary aim is used to set the sample size. In this case, it is not be necessary to have 80% power, but it certainly is nice if you do.

Number 3 above is used when you have a fixed number of observations and specify the power a-priori, and want to know what level of risk you could detect. This is useful when you can't specify a meaningful or minimum effect size. In this case you solve for minimum detectable effect size, and you have to then argue whether this would be meaningful if detected.

Some Important Issues

Effect Size: Effect size plays a key role in defining the power of your study. If you are doing an observational study and using power analyses to determine the number of observations, then you must first select an effect size that would be meaningful to detect. This is not always easy, and it must be done carefully and with deliberation. One approach is to consider "clinically meaningful" effects, such as differences that would lead to reduced morbidity or mortality. A second approach is to consider differences that would be meaningful at the population level, in which case small differences in common exposures may have important implications for public health. Lastly, you may be able to argue from other research, for example other studies with similar exposures or of the same disease, or from accepted (ideological) standards in your field of research. Clinical trials require not only arguing that the effect will be meaningful but that the intervention will be sufficiently strong to produce such an effect. **In all cases, your best defense is a pilot study, from which you justify expected effects. Your second best defense is to argue by analogy from earlier studies.**

Multiple Testing: Whether or not you must adjust your $\alpha$ error for multiple tests is a philosophical argument. If you have several specific aims, you have a dilemma. The most conservative approach is to divide the desired $\alpha$ error for a single test by the total number of planned tests, an approach that leads quickly to escalating sample sizes and budgetary nightmares.

There are several approaches to handling multiple tests, and which ever you choose must be justified clearly. Whatever you do, do not ignore the problem! At the very least, state that you are aware of it, but that because you have given *a prior* hypotheses you are protected against the worst offenses of multiple testing. This argument is more or less true, more if your hypotheses are orthogonal and less if they are closely related. Beware multiple testing with variables derived from the same measure (e.g., 144 nutrients and countless ratios from a food frequency questionnaire; weight, body mass index, and hip/waist ratios; or vitamins E, A, $B_1$, $B_2$, $B_6$ and $B_{12}$ from multivitamin supplements). There are other statistical solutions, and it is best to consult with a biostatistician who understands your research area. One approach is to specify a single test that must be significant before you complete subsequent tests, and there are accepted approaches that are not as costly as simply dividing your $\alpha$ error by the number of tests. Finally, consider moving as many tests as possible to secondary aims. Reviewers are not as strict about considering secondary aims as part of multiple tests for overall study power.

See examples of Power sections below.

**For a Case Control Study**

**Study Power and Minimal Detectable Odds Ratio**

Because the sample size of this study is fixed by the PCPT, we present power calculations in terms of the minimal detectable difference in the rate of biopsy proven prostate cancer between those in the upper quarter of fat intake (or any nutrient or food group intake), $p_1$, and the rate in lower quarter, $p_0$.

The difference was computed based on the standard power calculation formula for the difference between groups:[103]

$$\Delta = |p_1 - p_0| = \left( \frac{(Z_{\alpha/2} + Z_\beta)^2 \cdot 2\bar{p}(1-\bar{p})}{n} \right)^{1/2}$$

This formula yields the minimal detectable difference in outcome, $\Delta$, for a study with n subjects in the upper quartile of food intake and n subjects in the lower quartile. This difference can then be converted into a minimal detectable odds ratio, which can be interpreted as the minimal odds ratio for the upper quartile of intake versus lower quartile that will yield a 95% CI around the odds ratio that would exclude one.

The assumptions used in the computation were:

1. A two-sided significance level $\alpha$ equal to 5%.
2. Power (1-$\beta$) equal to 80%.
3. The average outcome, $\bar{p}$, in the combined upper and lower quartile of intake was assumed to be the 6%. This is based on the PCPT estimate that 6% of the control group who have biopsies will have detectable prostate cancer at seven years. This estimate is based on a variety of sources. First, the 10-year risk of *clinically overt* prostate cancer for a 65 year old white male is 4.3% and for a black male 6.1%.[104] In a prospective study of men negative for cancer at baseline (based on examination of tissue from TURP), there was a 6.3% incidence of prostate cancer by nine years (based on digital rectal examination alone for follow-up).[105] Because the PCPT will recruit men at higher risk (based on race and family history), will use an endpoint based on uniform biopsies and not screening, and will not eliminate participants at baseline with latent cancers detectable only by biopsy, the 6% estimate in the control group appears very conservative.
4. The number of men in each quartile of intake, n, was assumed to be 1,350. This was based on 9,000 men to be randomized into the PCPT control arm (proposal), with 60% (5,400) having complete data. The assumptions leading to the 40% loss are discussed above under "Endpoint Ascertainment and Final Sample Size." By definition, one quarter of the 5,400 men would fall into each quartile of intake.

Applying the above assumptions to the equation above yields d = .026. Translating this difference into $p_0$ and $p_1$ and an odds ratio (such that $(p_0 + p_1)/2 = .06$ and $p_1 - p_0 = .026$) yields $p_0$ = .047, $p_1$ = .073 and an odds ratio = 1.60. This suggests that the study will have 80% power to detect an odds ratio of prostate cancer of 1.6 for those in the upper vs. lower quartile of fat intake or saturated fat intake. This magnitude for the odds ratio is consistent with that observed in prior studies. Of the four case-control studies that had at least 250 cases and which computed total fat intake (Table 1), the relative risks of high versus low total fat consumption were 2.0, 1.5, 1.9, 0.8. Results for high-fat food groups and animal fat are similar.

*We also completed power analyses based on an expected 4% prevalence of more advanced, latent tumors. This yields d = 0.21, $p_0$ = .030, and $p_1$ = .051. We will have 80% power to detect an odds ratio of 1.75.*

There are three potential sources of additional statistical power. First, if there is a dose-response effect when we use all four quartiles of fat intake to calculate a test for trend, we will be able to detect an odds ratio of 1.54 for those in the upper quartile of intake vs. lower quartile. This is based on a test for linear trend in proportions.[106] Second, if we are able to combine the intervention with control groups, our sample size will double (although not our number of cases, if the intervention is effective). *Third, we can use an ordered, histopathologic scale to assess tumor grade rather than a binomial (presence/absence) outcome.*

For secondary aims a, b, and c, the power calculations are essentially identical except we assume that fruits, vegetables, fiber and some micronutrients may be protective. The minimal detectable odds ratio of prostate cancer for those in the upper quarter of nutrient intake versus lower quarter would be .62 (the inverse of 1.60). This magnitude of reduction (38% reduction in risk) is reasonably consistent with past studies (Table 2) and is appropriate as a degree of reduction which would have a potential public health impact.

For secondary aim d, comparing the effects of finasteride in participants with low- vs. high-fat intakes, the power to detect subtle differences is low. However, there is fair power to detect sizable interactions. One example of such an interaction is that there will be no intervention effect in the low-fat diet group. Keeping assumptions about the overall effects of finasteride (25% reduction in incidence) and the effects of dietary fat (RR=1.6 in lowest versus highest quartiles), this would require detecting a difference between the effects of finasteride with a low fat diet (0% reduction) versus a high fat diet (40% reduction). Based on a linear trend test comparing slopes, there would be 70% power with two-sided alpha error of 5% to detect such an interaction.

---

## Cohort Study

**Expected numbers**. After discounting rates by 20% to account for a possible "healthy volunteer effect", we expect over 2300 cancer cases to be diagnosed in the cohort over the 2.25 year follow-up period (Table 10). Although total cancer incidence is the primary endpoint, cancers will be grouped into major SEER groupings, as shown in Table 10. The other assumptions on which these numbers are based and a discussion of the detectable risk ratios are given in Section D.9. This table also presents the number of endpoints that are expected if this cohort were followed-up in the future.

**Table 10. Expected Number of Events and Detectable Risk Ratios**

| | **Expected N for follow-up period of:** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Outcome** | 2.25 Years | | | 6.25 Years | | | 10.25 Years |
| | Total | Men | Women | Total | Men | Women | Total |
| Total cancer cases (invasive) | 2360[c,e] | 1310[c,e] | 1050[c,e] | 6360[c,e] | 3520[c,e] | 2840[c,e] | 9910[c,e] |
| Lung | 420[b,e] | 240[a,d] | 180[a,d] | 1170[c,e] | 670[c,e] | 500[b,e] | 1860[c,e] |
| Prostate | - | 510[b,e] | - | - | 1420[c,e] | - | 2270[c,e] |
| Breast | - | - | 340[b,e] | - | - | 910[c,e] | 1420[c,e] |
| Colorectal | 230[a,d] | 130 | 100 | 670[c,e] | 360[b,d] | 310[b,e] | 1120[c,e] |
| Colon | 180[a,d] | 100 | 80 | 530[c,e] | 280[b,d] | 250[b,d] | 890[c,e] |
| Rectal | 50 | 30 | 20 | 140[a,d] | 80 | 60 | 240[b,d] |
| Corpus Uteri | - | - | 80 | - | - | 220[a,d] | 340[b,e] |
| Bladder | 50 | 40 | 10 | 140[a,d] | 110 | 30 | 240[b,d] |
| Non-Hodgkins Lymphoma | 50 | 25 | 25 | 140[a,d] | 75 | 65 | 230[b,d] |
| Oral | 60 | 40 | 20 | 170[a,d] | 110 | 60 | 270[b,d] |
| Melanoma | 60 | 35 | 25 | 160[a,d] | 90 | 70 | 240[b,d] |
| Pancreas | 50 | 25 | 25 | 150[a,d] | 75 | 75 | 250[b,d] |
| Total deaths | 2910[c,e] | 1730[c,e] | 1180[c,e] | 8690[c,e] | 5090[c,e] | 3600[c,e] | 15130[c,e] |

[a-e] Detectable risk ratio (RR) based on trend test:

| | RR for one category increment | Non-users of supplement | Users of supplement (tertiles) lowest | middle | highest | | RR for one category increment | Non-users of supplement | Users of supplement (tertiles) lowest | middle | highest |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.80 | 1.0 | 0.80 | 0.64 | 0.51 | or | 1.25 | 1.0 | 1.25 | 1.56 | 1.95 |
| B | 0.85 | 1.0 | 0.85 | 0.72 | 0.61 | or | 1.18 | 1.0 | 1.18 | 1.38 | 1.63 |
| C | 0.90 | 1.0 | 0.90 | 0.81 | 0.73 | or | 1.11 | 1.0 | 1.11 | 1.24 | 1.37 |
| *D* | *0.70* | -- | *1.00* | *0.70* | *0.49* | *or* | *1.43* | -- | *1.00* | *1.43* | *2.04* |
| *E* | *0.80* | -- | *1.00* | *0.80* | *0.64* | *or* | *1.25* | -- | *1.00* | *1.25* | *1.56* |

Two subscripts are given: the first is for vitamins E, C and calcium *for the 4 level comparison and the second for the 3 level comparison (users only).*.
No subscript indicates detectable RR < 0.80.

**Study power.** The projected number of events and detectable risk ratios are given in Table 10. These were computed from the following assumptions: 1) 40,000 women and 35,000 men; 2) age distribution in the study population matching the age distribution in the Supplement Pilot Study: women 26%, 19%, 15%, 20%, and 20% and men 26%, 19%, 18%, 17%, and 20% at ages 50-54, 55-59, 60-64, 65-69, and 70-74, respectively; 3) vitamin supplement use in the study population matching the use reported in the pilot study (Table 5); 4) one-quarter of the study population recruited at each of 1 year, 1½ years, 2 years, and 2½ years after the project begins; 5) cancer incidence rates by age and sex at 80% of the corresponding rates from the western Washington SEER registry for years 1990-95; the use of 80% assumes a healthy volunteer effect; 6) death rates by age and sex at 60% of the U.S. vital statistics rates for whites for years 1991-93 [133]; the 60% is for a healthy volunteer effect and is based on our experience in CARET (unpublished data); 7) 2% per year loss of participants to follow-up (i.e., their vital status and cancer status cannot be determined) expected to be conservative based on the discussion in Section D.8; 8) primary analysis is a test for trend among four categories of supplement use: none and three tertiles of level of use among users, using all events occurring post-enrollment with the true risk ratio constant between successive categories; and 9) two-sided significance level of 0.05 and power of 80%.

Table 10 (see footnotes) gives the risk ratios detectable for the total cancer outcome at the primary analysis at the end of the 5 year study (mean of approximately 2¼ years of follow-up since recruitment). For men and women *considered separately*, we will be able to detect a *27%* reduction in risk (RR=*.73* for highest third of supplement use vs. no use of that supplement) or a *37*% increase in risk (RR=*1.37* among highest third of users) for each of the four types of supplements of interest (vitamin C, vitamin E, calcium and multivitamins plus minerals). This level of risk or benefit is consistent with prior studies *(4a)* and would be of public health significance. *We also have power to detect a RR=.64 for highest vs. lowest third of use for the trend test limited to users.* Specific cancer site relative risks detectable at 2¼, 6¼ and 10¼ years of follow-up (if future applications were funded) are also given. By 6¼ years follow-up, there would be sufficient power to detect a 27% risk reduction (RR=.73) for lung, prostate, breast and colorectal cancer. *Also, if we have at least 35% success in obtaining DNA from participants, we should have a reasonable number of events in this group by 6 ¼ years. The number of expected events by 6 ¼ years in the group with usable DNA would be approximately the number listed under 2 ¼ years in Table 10 (for the full sample).*

An advantage of the planned design is the anticipated high level of supplement intake in the study population. If the true relation between relative risk and dosage of supplement is log-linear (i.e., constant relative risk for equal increments in dosage of supplement), and if for a given study population the detectable relative risk per unit change in dosage is *r*, then increasing the standard deviation of dosages of supplements in the study population by a factor *f* changes the detectable relative risk per unit change in dosage to approximately $r^{1/f}$ (131). Thus, a 50% increase in standard deviation of dosages would change detectable relative risks of 0.95, 0.90, 0.85, and 0.80 to 0.97, 0.93, 0.90, and 0.86, respectively, and a doubling would change them to 0.97, 0.95, 0.92, and 0.89. We have shown that the increase in power is even greater in the presence of measurement error (85). Thus, by recruiting a study population with a wide range of supplement intake, we increase the power of the study to detect health effects of public health significance.

---

**For a Randomized Trial**

**STATISTICAL POWER**

Table 11 gives the summary of study power for the principal endpoints and the secondary endpoints related to dietary change at the one year follow-up. The criteria for power calculations are based on two-sided tests with alpha error of five percent. We based our power analyses on a minimum difference in percentage of energy from fat of two percentage points between UC and PSH; for fiber the minimum difference is two grams. Choosing these minimum effect sizes for this intervention trial has been difficult, and we explain our reasoning in detail below.

We believe that the intervention effect sizes of two percentage points in energy from fat and two grams of fiber per 1000 Kcal are reasonable targets for our Personalized Self-Help intervention, for several reasons. First, these effects sizes are scientifically important. While modest compared to intensive, clinical interventions that are designed to treat underlying disease, these effects sizes are meaningful public health goals for population-level disease prevention. *We point out to reviewers that a two percentage point reduction in percentage of energy from fat is a 5.6% reduction in total fat intake (based on a 2000 Kcal diet at 36% energy from fat).* Second, we expect larger effects sizes then those from our earlier studies, because we will add components to enhance the potency of our interventions, especially those related to use of fruits and vegetables. Third, the effects of self-help interventions tend to be cumulative, both in smoking cessation programs[61] and our previous Primary Care study (Table 1). Thus, our estimates of effectiveness at 12 months may underestimate the intervention's *long-term impact*. Fourth, we place this research in the context of community interventions, in which self-help interventions would be a component of a comprehensive intervention that targets food services, supermarkets, media and the health care industry. We would expect a greater impact of the PSH intervention when coupled with interventions that target environmental determinants of dietary behavior. In summary, the above arguments support more realistic goals for public health dietary change interventions. Achieving the effects proposed for this intervention in a single year would represent significant movement toward the Year 2000 goals.

**Table 11. Power Analyses ($\alpha_2$=.05) for Principal and Secondary Endpoint Measures with 620 Participants per Treatment Arm**

| Instrument | Measure | Smallest Meaningful Intervention Effect Between Each Arm | Power (1-β) | Minimum Detectable Difference with 80% Power |
|---|---|---|---|---|
| **Principal Endpoints** | | | | |
| 24-Hour Diet Recall | Fat (% En) | 2 percentage points | .80 | 2.0 |
| | Fiber (g/1000 Kcal) | 2 g | .88 | 1.8 |
| **Secondary Endpoints** | | | | |
| Fat and Fiber Behavior (FFB) | Fat Scale[1] | 0.14 units | > .95 | .057 |
| | Fiber Scale[1] | 0.19 units | > .95 | .064 |
| Stage of Change | Percent moving into action stage[2] | Fat: 18 percentage points | .94 | .14 |
| | | Fiber: 17 percentage points | > .95 | .11 |

[1]Scales range from 1 to 4, and are not directly comparable to those on Tables 1 and 2.

[2]Precontemplation, contemplation, decision at baseline moving to action or maintenance at follow-up

We used data from both published studies[62] and our own datasets to estimate the variances of nutrients measured by 24-hour recall. We used data from the Dietary Intervention Trial in Primary Care Practices study to transform our effect sizes expressed as nutrients into the metric of the fat- and fiber-habits scales: A unit change in the fat-related habits scale corresponds to a 14 percentage point change in fat (%en) and a unit change in the fiber-related habits scale corresponds to a 5 g/1000 Kcal change in fiber. Finally, for stage of dietary change, we have three studies in which participants completed the two stage of change questionnaires separated by between 3 and 12 months. On the basis of these studies, we estimate that approximately 28% of participants will be in a pre-action stage for fat and 46% will be in a pre-action stage for fiber, and that intervention will increase the percentage of participants moving into an action stage by approximately 18% for fat and 12% for fiber.

Based on the above, our design is to have 620 participants in each treatment arm active throughout the study. We propose to recruit 730 per arm to allow for a 15% loss to follow-up over 12 months. This will give us 80% power to detect minimal meaningful differences in the principal endpoint measures, and 90% or greater power to detect differences based on the FFB and movement through stages of change. *For the 18 month follow-up, in which there will be 310 participants per arm, we will have greater than 95% power to detect the minimal intervention effect using the FFB. Power to detect shifts through stages of change are somewhat less, 70% for fat and 83% for fiber. Though the power to detect differences based on the FFB at 18-*

*months is large, we chose to keep this sample size to support analyses of factors, for example gender or age, related to long-term intervention effectiveness.*

The power calculations for the principal endpoints, differences in fat (%en) and fiber intake from 24-hour recalls, are based on two, very conservative assumptions. First, we assumed that there is no autocorrelation between individual assessments of diet between baseline and one year (due to the high intra-individual variability of 24-hour recall estimates of nutrient intake). We observe correlations between repeated 24-hour recall measures of fat that are less than 0.2, and this level of autocorrelation would improve study power only marginally. Second, we considered an evaluation based on a post-test measure alone. Such an evaluation would have more power, because the variance of the difference in mean nutrient intake between intervention and control arms at the 12-month follow-up will be less than the variance of the differences in mean change from baseline to 12 month follow-up. However, an evaluation based on post-test measures only would require that mean nutrient intake at baseline is well balanced between the two arms. We do not feel comfortable relying upon that assumption.

---

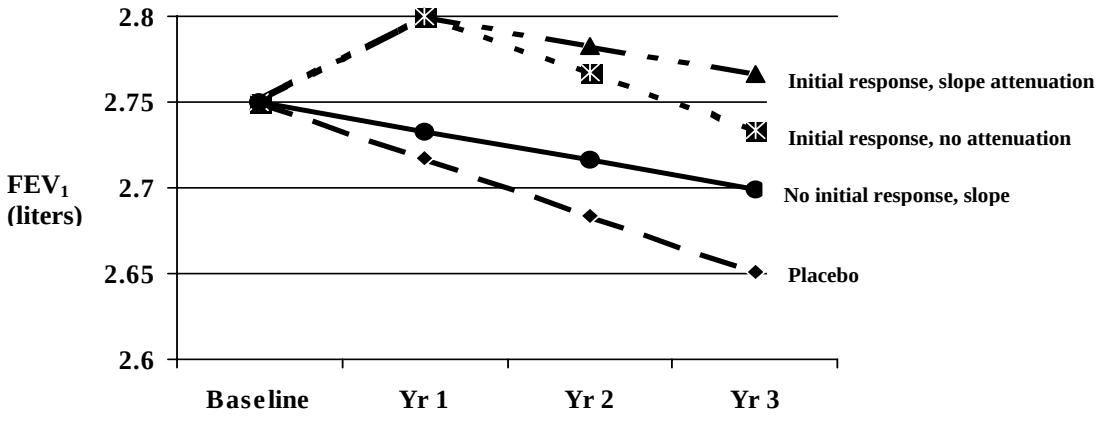
**For an Ancillary Study to a Randomized Trial**

**Respiratory**

**Ancillary Study: Design assumptions, sample size calculation and study power**

The starting point for all design considerations is to estimate the expected response to treatment. There are three possible patterns for the intervention effect expected in this study. Treatment with nutritional antioxidant supplements may produce an initial response in $FEV_1$ (expected to be an increase in *level*), a change in the slope of decline in $FEV_1$ (expected to be an attenuation in *slope*), or a combination of both. The size of the intervention effect will depend on the pattern of response to treatment, and therefore on the magnitude of the increase in $FEV_1$ and/or the degree of attenuation in the slope of $FEV_1$ decline. Figure 1 shows various scenarios for the intervention effect according to this scheme, and the placebo line shows the expected average decline in $FEV_1$ over three years in untreated participants. As shown in the figure, the magnitude of the treatment effect is a function of *both* the initial response and the attenuation in slope. Thus, while we describe our sample size calculations in terms of the 3-year intervention effect, we recognize and describe below information related to these different scenarios for the pattern of the treatment effect.

**Figure 1.** The intervention effect on $FEV_1$ may take one of 3 patterns, as shown. The **placebo** line shows the expected average yearly decline in $FEV_1$ among untreated men.



The following estimates of five parameters were used in calculating sample size:

Decline in $FEV_1$ in the placebo group: During mid- to late adulthood, $FEV_1$ declines on average about 25 ml/year, with a steeper decline among cigarette smokers of about 75 ml/year (16). We assumed an average rate of decline in the placebo group of 33 ml/year, given a mix of about 85% nonsmokers and 15% smokers. This estimate of expected change in 3 years in the placebo is conservative for two reasons: (1) our estimate for annual rate of decline may be low, indeed recent data over 30 years of follow-up in the Seven Countries Study suggests the annual rate of decline in nonsmokers is about 38 ml/year (25), and (2) if the mix includes more smokers, then the actual average rate of decline in the placebo group may be greater still. Thus, we expect that after three years of follow-up there will be a -100 ml change in $FEV_1$ in the placebo group.

Standard deviation of the change in $FEV_1$ from baseline to follow-up: The standard deviation of $FEV_1$ is estimated to be about 680 ml (69, 99). The correlation in repeated measures of $FEV_1$ over 3 years is estimated to be 0.80 (69). The variance of the change in $FEV_1$ i.e., the variance of the difference, $\sigma_d^2$, is equal to (2 x variance of single measure) x (1- intraclass correlation coefficient). Using the above information, the standard deviation of the change in $FEV_1$ was estimated as 430 ml.

Loss to follow-up (drop outs, deaths): The loss to follow-up and non-compliance is conservatively estimated at 2.5% annually. We expect the actual losses to be less given that men enrolled in this study are highly motivated volunteers.

Decline in $FEV_1$ in group treated with selenium alone: We used three lines of evidence to estimate the expected effect of the selenium intervention. First, we considered the cross-sectional association of selenium with $FEV_1$. Second, we considered observational longitudinal epidemiologic studies of diet and $FEV_1$. Finally, we considered the longitudinal Lung Health Study (99) and the effect size produced by the smoking cessation intervention in that study. A cross-sectional study found that serum selenium was strongly associated with $FEV_1$ (1): a standard deviation increase in serum selenium (17 ng/ml) was associated with an increase of 25 ml in $FEV_1$. Given that selenium supplementation (200 µg/day) is expected to raise mean serum selenium by 67% (65), i.e., about 5 standard deviations, supplementation could lead to a difference of 125 ml in $FEV_1$ in treated vs. untreated men based on the magnitude of the cross-sectional effect. A difference this large is not expected as some of the cross-sectional effect likely resulted from long-standing differences over time. In a longitudinal study of adults, the sub-group with unfavorable dietary change (decreased fruit intake) had a steeper annual decline in $FEV_1$ compared to those with no change in diet (37). Over three years of follow-up the change in $FEV_1$ was –330 ml and –9 ml, respectively, in the two groups, yielding a difference of 321 ml. Once again, we do not expect a difference this large: indeed it is unclear what proportion of this difference is due to diet change *per se* versus other behavioral changes that may accompany unfavorable dietary change. In the Lung Health Study (99) the difference in $FEV_1$ between sustained quitters and continuing smokers was about 250 ml over 3 years of follow-up. This difference was the result of initial improvement in $FEV_1$ and a subsequent attenuation in rate of decline (as per top line in figure 1), and is presumably due to the change in oxidative burden brought about by smoking cessation. We reasoned that the effect of selenium intervention would not be as great as the effect achieved by successful smoking cessation.

The selenium intervention proposed herein is expected to contribute to an improvement in the antioxidant defenses in the lung, thus the smoking intervention, the dietary change, and the selenium effect sizes described above are relevant as they presumably result from similar changes in the antioxidant "shield". Based on the three lines of evidence above, we powered this study *conservatively* to detect a difference in $FEV_1$ of 85 ml over 3 years, or slightly more than half the cross-sectional effect.

Decline in $FEV_1$ in group treated with vitamin E alone: To estimate this effect, we compared the cross-sectional data on vitamin E to that available for selenium. In a cross-sectional study of serum vitamin E and $FEV_1$ (1), a standard deviation increase in serum vitamin E (11.5 µmol/l)

was associated with an increase of 40 ml in $FEV_1$. Vitamin E supplementation (400 mg/day) is expected to raise serum vitamin E by 15 µmol/l or 1.4 standard deviations, and thus may increase $FEV_1$ as much as 50 ml. Thus, the effect size for Vitamin E was estimated to be about half of the effect size for selenium, given the comparison in their cross-sectional effects (50/125=40%). After 3 years of follow-up, we hypothesize that the effect of vitamin E intervention will be about half the effect of the selenium intervention. Following this argument, we expect about a 57 ml decrease in $FEV_1$ in the vitamin E group, as compared to a 100 ml decrease in $FEV_1$ in the placebo group.

Decline in $FEV_1$ in group treated with both selenium and vitamin E: In the cross-sectional study (1), the effects of vitamin E and selenium were independent. We therefore assumed that the effects of the study supplements will be additive. Based on this assumption, after 3 years of follow-up an increase of 28 ml in $FEV_1$ is expected in the group receiving *both* selenium and vitamin E, to yield an overall difference with the placebo group of +28 ml (i.e., treatment effects are additive: Se effect is 85 ml, vitamin E effect is 43 ml; combined effect is 128 ml).

Sample size calculation: We used the formula below to compute the sample size under the following conditions: expected treatment effect of 85 ml (=d*), ß=0.10, $\alpha$=0.018 (see above on multiple comparisons), $\sigma_d^2$ = 430 ml, and r=1.

$$N = (Z_{\alpha/2} + Z_{\beta})^2 \times \sigma_d^2 (r+1)/ (d^*)^2 r$$

The required number of men, N, in each treatment group is 700 at year 3 when study analyses are completed. Allowing for an annual attrition of about 2.5%, we will enroll 750 men per cell. The total sample size is 3,000 participants, to be recruited from 20 study sites in order to complete recruitment within a one-year timeframe (about 150 per site).

This sample size calculation, which provides us with adequate power for the primary aims of the proposed study, is the basis for our targeted enrollment. We also calculated the power of the study for all other outcomes (Table 2). We also examined how changes in our assumptions would affect power. For example, we assumed an annual attrition of 2.5%, but if the attrition was 5%, we would have 88% power for specific aim 1.1 after three years of follow-up. Because the vitamin E effect is expected to be about half the for selenium effect, longer follow-up will be needed to detect this effect. We estimate adequate power by year 6, and plan to use interim results at year 3 for Vitamin E to motivate a continuation proposal. Specific aim 2.3 proposes effect modification, whereby the effect of supplementation is hypothesized to vary across levels of oxidative stress. Given the following three conditions, we have 90% power at year 3 to detect an interaction: (1) total group is split at the median of the urinary oxidative stress biomarker, (2) treatment has no effect in low oxidative stress ($FEV_1$ decline similar to placebo), (3) treatment causes an overall *improvement* in $FEV_1$ of 70 ml in 3 years in high oxidative stress group.

**Table 2. Summary of power calculations†**

| Comparison | Intervention effect§ $((FEV_1)_b-(FEV_1)_f)_I-((FEV_1)_b-(FEV_1)_f)_P$ | Power |
|---|---|---|
| ***3 years follow-up*** | | |
| Selenium vs. placebo | 85 ml | 90% |
| Selenium & Vitamin E vs. placebo | 128 ml | >99% |
| Vitamin E* vs. placebo | 43 ml | 38% |
| Combined vs. Vitamin E | 85 ml | 90% |
| Combined vs. Selenium | 43 ml | 38% |
| ***6 years follow-up*** | | |
| Vitamin E* vs. placebo | 85 ml | 88% |
| Combined vs. Vitamin E | 170 ml | >99% |
| Combined vs. Selenium | 86 ml | 89% |

†all calculations based on annual attrition of 2.5% in each cell

§ subscripts *b* and *f* refer to baseline and follow-up, and the subscripts *I* and *P* refer to intervention and placebo groups

*expected effect size for Vitamin E is predicted to be about half the effect of selenium

We also considered what level of difference between treatment and non-treatment groups would be clinically meaningful. If the difference in the rate of decline was as small as 15ml/year without an initial increase in $FEV_1$, similar to the overall effect in the Lung Health Study (99), then the annual decline in $FEV_1$ in the treatment group would be about 18 ml/year (expected decline in absence of treatment =33 ml/year – 15 = 18), i.e., a 45% attenuation in the annual rate of decline. Given this effect size, the trial would need longer follow-up. At three years the magnitude of the effect will be evident, and if it is at this level or greater, we plan to seek funding to continue the study to assess whether the attenuation in decline is sustained with continued supplementation. At 6 years, with 2.5% annual attrition yielding a final estimate of 550 per cell, there is 88% power to detect a difference of 90 ml in $FEV_1$ (i.e., the expected difference in 6 years if supplementation produces a 45% attenuation in the slope of $FEV_1$ decline).